OVERVIEW

- **TEXT MINING INVOLVEMENT HISTORY**
- TEXT MINING DEFINITIONS
- GOALS/ OBJECTIVES/ APPLICATIONS
- TEXT MINING COMPONENTS
- BARRIERS TO TEXT MINING IMPLEMENTATION
- PILOT TEXT MINING PROGRAM
- LESSONS LEARNED FROM PILOT PROGRAM
- NEXT STEPS

TEXT MINING INVOLVEMENT HISTORY

- PURPOSE:
- DEVELOP TEXT MINING TO SUPPORT PROGRAM OFFICERS

- THREE DISTINCT PHASES
  - PRE-PHASE 1
    - 1991-1997 (PART-TIME) ~300K TOTAL
  - PHASE 1
    - 1998 (FULL-TIME) ~150K TOTAL
  - POST-PHASE 1
    - 1999-2000 (PART-TIME) ~50K TOTAL
    - NON-CORPORATE FUNDING

THREE PHASE SUMMARIES

- PRE-PHASE 1
  - DEVELOP FULL-TEXT MINING TO SUPPORT S&T
  - GAIN CREDIBILITY, VISIBILITY
- PHASE 1
  - ENHANCE ROLE OF TECHNICAL EXPERTS IN STUDIES
  - EXAMINE DIFFERENT DATABASES

THREE PHASE SUMMARIES (CONT'D)

- POST-PHASE 1
  - DEVELOP BETTER UNDERSTANDING OF S&T TEXT MINING
    - HIGH QUALITY REQUIREMENTS
    - SCOPE OF APPLICATIONS
  - LATEST WORK ON INFORMATION RETRIEVAL, TEXT MINING, LITERATURE-BASED DISCOVERY, CITATIONS MOST EXCITING
  - CANNOT DISCUSS UNTIL PATENT APPLICATIONS FILED, PAPERS ACCEPTED FOR PUBLICATION

IMPACT OF INVOLVEMENT

- DEVELOPED FULL TEXT CO-WORD TEXT MINING FOR S&T EVALUATION
  - PREVIOUS EFFORTS USED KEY WORDS ONLY
- PUBLICATIONS
  - 16 PAPERS IN PEER REVIEWED JOURNALS
  - 9 PAPERS IN PEER REVIEWED CONF. PROCEED.
  - 1 BOOK CHAPTER
  - 2 PAPERS ON WEB SITES
  - 4 PAPERS SUBMITTED TO JOURNALS
  - 10 PAPERS TO BE SUBMITTED TO JOURNALS
- JOURNALS
  - JASIS, IPM, JIS (INF TECH)
  - CHEMICAL REVIEWS, JOURNAL OF AIRCRAFT, ANALYTICAL CHEMISTRY (NON-INF TECH)

OVERVIEW

- TEXT MINING INVOLVEMENT HISTORY
- TEXT MINING DEFINITIONS
- **GOALS/ OBJECTIVES/ APPLICATIONS**
- TEXT MINING COMPONENTS
- BARRIERS TO TEXT MINING IMPLEMENTATION
- PILOT TEXT MINING PROGRAM
- LESSONS LEARNED FROM PILOT PROGRAM
- NEXT STEPS

DATA MINING GOALS/ OBJECTIVES

- **DEVELOP CAPABILITY TO ALLOW**:
  1) PROGRAM OFFICERS;
  2) SENIOR MANAGEMENT;
  3) IFO;
  4) NSAP;
  5) NRL RESEARCHERS;
  6) WARFARE CENTER/ TRANSITION AGENTS;
  7) PROGRAM REVIEWERS; OTHERS

  FULL ACCESS AND INSIGHT TO RELEVANT GLOBAL S&T DATA TO SUPPORT:
  1) DISCOVERING AND INNOVATING,
  2) PLANNING AND EXECUTING,
  3) MANAGING AND TRANSITIONING,

  OF THE ONR S&T PROGRAM

DATA MINING GOALS/ OBJECTIVES (CONT'D)

- **HELP ANSWER FOLLOWING GENERIC QUESTIONS:**
  - WHAT S&T IS BEING DONE GLOBALLY?
  - WHO IS DOING IT?
  - WHERE IS IT BEING DONE?
  - WHAT MESSAGES CAN BE EXTRACTED FROM GLOBAL S&T?
  - WHAT IS NOT BEING DONE?
  - **--->**WHAT SHOULD WE BE DOING DIFFERENTLY?

OVERVIEW

- TEXT MINING INVOLVEMENT HISTORY
- TEXT MINING DEFINITIONS
- GOALS/ OBJECTIVES/ APPLICATIONS
- TEXT MINING COMPONENTS
- BARRIERS TO TEXT MINING IMPLEMENTATION
- **PILOT TEXT MINING PROGRAM**
- LESSONS LEARNED FROM PILOT PROGRAM
- NEXT STEPS

CONTENTS

- WILL PRESENT TECHNICAL RESULTS FROM RECENT STUDIES; MAINLY AIRCRAFT
- WILL PRESENT IMPLICATIONS FOR FUTURE STUDIES
- WILL PRESENT TECHNICAL RECOMMENDATIONS FOR FUTURE STUDIES

RECENT STUDIES

- **PURPOSE :**
  - DEMONSTRATE FEASIBILITY AND ADDED VALUE OF EMPLOYING TOPICAL AREA EXPERTS
  - UNDERSTAND HOW TO APPLY TEXTUAL DATA MINING TO A BROAD SPECTRUM OF DATABASES

- **STRUCTURE**
  - CONTAINS THREE COMPONENTS OF PRIOR ACTIVITIES
    1) ITERATIVE INFORMATION RETRIEVAL FROM DIFFERENT DATABASES
    2) INFORMATION PROCESSING
       BIBLIOMETRIC STUDIES OF RETRIEVED RECORDS
       COMPUTATIONAL LINGUISTICS STUDIES OF RETRIEVED RECORDS
    3) INFORMATION INTEGRATION
       INTERPRETATION AND ANALYSIS OF RETRIEVED RECORDS AND COMPUTER OUTPUT

RECENT STUDIES (CONT'D)

- **THREE STUDIES COMPLETED FROM FY98 PROGRAM**
  - SHIP HYDRODYNAMICS (SINGLE TECHNOLOGY/ RESEARCH AREA)
  - AIRCRAFT SCIENCE AND TECHNOLOGY (MULTI-TECHNOLOGY SYSTEM)
  - FULLERENES (SINGLE RESEARCH AREA)

- **TANGIBLE OUTPUTS INCLUDE**:
  1) MULTIPLE RELEVANT RECORDS
  2) REPORT OF GLOBAL ACTIVITY IN TOPICAL S&T AREA
  3) JOURNAL PAPER FOR EACH TOPICAL AREA.

RESULTS FROM RECENT STUDIES

- **AIRCRAFT FINDINGS**

  - WILL SUMMARIZE CONTRACTOR PRESENTATION
  - INTERSPERSED VUGRAPHS FOR CONTEXT
  - -INCLUDE RESULTS FROM OTHER STUDIES WHERE USEFUL

QUERY - RESULTS FROM RECENT STUDIES

EXAMPLES OF OUTPUT

- **INPUT QUERY/ COMPREHENSIVE DATABASE OF RELEVANT RECORDS**

  - START WITH SOME INITIAL QUERY
  - DIVIDE RECORDS RETRIEVED INTO TWO GROUPS; RELEVANT AND NON-RELEVANT RECORDS
  - USE PHRASE FREQUENCY AND PHRASE PROXIMITY ALGORITHMS TO OBTAIN PHRASE FREQUENCY AND PHRASE RELATIONSHIP PATTERNS CHARACTERISTIC OF EACH GROUP
  - ADD PHRASES CHARACTERISTIC OF RELEVANT GROUP TO QUERY; SUBTRACT PHRASES ("NOT" BOOLEAN) CHARACTERISTIC OF NON-RELEVANT GROUP FROM QUERY
  - ITERATE UNTIL CONVERGENCE OBTAINED
  - ***MOST CRITICAL PART OF TEXT MINING***

RESULTS FROM RECENT STUDIES (CONT'D)

SAMPLE QUERY - SHIP HYDRODYNAMICS

- (hydrodynamic* or hydromechanic* or fluid flow or potential flow or incompressible flow or wake or turbulen* or vort*) AND (bound* or ship* or surface* or hull* or fish or dolphin) NOT (accret* or adhes* or adsor* or aggregat* or bacter* or bear* or black hole or carbon* or cluster* or colli* or colloid* or combustion or crystal* or dissol* or emiss* or erosion or flame* or fractur* or gala* or grain* or ion* or larva* or lubrica* or melt* or membrane* or microscop* or mineral* or molecul* or organ* or permea* or plasm* or poro* or protein* or rock* or sediment* or shell* or shock or star or stars or stellar or sulf* or surface brightness or weld* or x-ray ageostrophic or animal* or antarctic or arctic or bay or bio* or cancer or CFC* or cilia or climat* or cloud* or coloni*or cosm* or crack* or cultivation or cumulus or diatom* or DNA or dunes or earthquake* or eco* or fermi or fluidised bed* fluidized bed* or greenhouse or gyre* or hydrographic or intertidal or Josephson or leaf or liposome* or monsoon* or muddy or nucl* or nutrient* or ozone or photolysis or phytoplankton or quantum or Rossby or sand or snow or soil or strato* or superconduct* or tropopause or undercurrent or ventricular or volcan* or zoo* or ablation or agglomeration or algal or alto* or astro-physics or astronomy or Benard convection or baroclinic* or barotropic* or blood flow or botan* or Brownian motion or capillary or cardiolog* or carotid or casting or CCD or cells or computational combustion dynamics or condensation or cyclon* or Darcy* or deep drawing or deposition or drainage or dredg* or drying or Ekman or electrochem* or environment*or enzyme* or estuary flow or fault* or film or foundry or fractal* or geostrophic or glycolipid* or granular or groundwater or Gulf-stream or heart or hydrology or hypersonic or ice mechanics or insect or irrigation or Kelvin-Helmholtz or laser welding or lipid* or liquid metal* or liquid-metal or locomotion or mantle or manufact* or materials or medical or microgravity or micromolecular or microscale or mining or molding or molten or Oseen or osmosis or physiolog* or pollution or polyphase flow or powder or preditor* or protozoa or pylori* or rain* or rarefied gas or reacting flow* or refuse or resuspension or roller* or rolling or scour* or seals or seismic or siltation or sintering or slag or solar or soldering or solenoid* or solidification or storm or sun or superfluid or supersonic or suspension* or tecton* or tide* or tidal or tokamak or tribology or turbidity or ultrasonic* or upwelling)

AIRCRAFT - DATABASES

- **SCIENCE CITATION INDEX**
  - **APPROXIMATELY 5600 JOURNALS & MAGAZINES**
  - **PHYSICAL, ENGINEERING & LIFE SCIENCES BASIC RESEARCH.**
  - **1991 - MID 1998.**
  - **PRODUCED 4346 APPLICABLE RECORDS.**

- **ENGINEERING COMPENDEX**
  - **APPROXIMATELY 2600 JOURNALS & CONFERENCE PROCEEDINGS.**
  - **MAINLY APPLIED RESEARCH AND TECHNOLOGY**
  - **1990 - MID 1998**
  - **PRODUCED 15,673 APPLICABLE RECORDS.**

AIRCRAFT -

DATABASE DEVELOPMENT -OBSERVATIONS

- **SCI**

  - REQUIRED SIGNIFICANT EFFORT TO DEVELOP QUERY FOR COMPREHENSIVE HIGH S/N. RELEVANT RECORDS
  REQUIRED A QUERY THAT CONSISTED OF 207 TERMS.

>>>> START WITH “AIRCRAFT” ; SUBTRACT NON-RELEVANT TERMS<<<<

- **EC**

  - CONSIDERABLY MORE FOCUSED ON JOURNALS/PUBLICATIONS OF
  INTEREST. VERY FEW EXTRANEOUS RECORDS GENERATED
  WITH 13 TERM QUERY.

*****

QUERY - LESSONS LEARNED FROM RECENT STUDIES

- **VALUE OF ITERATIVE QUERY APPROACH**

  - ALLOWS INCREASED RATIO OF RELEVANT/ NON-RELEVANT RECORDS; HIGHER SIGNAL-TO-NOISE RATIO
    - NOISE REDUCTION LESS IMPORTANT FOR SMALL RETRIEVALS
    - NOISE REDUCTION VERY IMPORTANT FOR LARGE RETRIEVALS
    - IMPROVES ANALYSIS RESULTS - KET LAW

  - ALLOWS MORE RECORDS IN FOCUSED FIELD TO BE RETRIEVED; INCREASED SIGNAL
    - USES LANGUAGE OF AUTHORS

  - ALLOWS MORE RECORDS IN ALLIED FIELDS TO BE RETRIEVED

  - ALLOWS POTENTIALLY RELEVANT RECORDS IN DISPARATE FIELDS TO BE RETRIEVED

BIBLIOMETRICS -RESULTS FROM RECENT STUDIES

EXAMPLES OF OUTPUT

- **BIBLIOMETRICS**

  - PROLIFIC AUTHORS
  - JOURNALS CONTAINING RELEVANT PAPERS
  - ORGANIZATIONS PRODUCING RELEVANT PAPERS
  - COUNTRIES PRODUCING RELEVANT PAPERS
  - MOST CITED AUTHORS
  - MOST CITED PAPERS
  - MOST CITED JOURNALS

RESULTS FROM RECENT STUDIES (CONT'D)

- **BIBLIOMETRICS - MOST CITED AUTHORS - AIRCRAFT**
  (CITED BY OTHER PAPERS IN DATABASE**)**

  - ERICSSON-LE,117
  - JOHNSON-W,97
  - MIELE-A,96
  - DOYLE-JC,82
  - TISCHLER-MB,80
  - SRINIVASAN-GR,78
  - PETERS-DA,75
  - HODGES-DH,70
  - HESS-RA,60
  - FRIEDMANN-PP,55
  - CHATTOPADHYAY-A,55
  - NEWMAN-JC,54
  - FARASSAT-F,53
  - JAMESON-A,50
  - MENON-PKA,50

RESULTS FROM RECENT STUDIES (CONT'D)

- **BIBLIOMETRICS - MOST CITED AUTHORS - FULLERENES**

  - KROTO HW,4328
  - KRATSCHMER W,3472
  - IIJIMA S,1787
  - TAYLOR R,1721
  - HADDON RC,1711
  - HEBARD AF,1563
  - DIEDERICH F,1476
  - FOWLER PW,1469
  - BETHUNE DS,1466
  - HIRSCH A,1264
  - EBBESEN TW,1145
  - ALLEMAND PM,1103
  - HEINEY PA,1064
  - HAUFLER RE,1021

RESULTS FROM RECENT STUDIES (CONT'D)

- **BIBLIOMETRICS - MOST CITED PAPERS - AIRCRAFT**

  - 'JOHNSON-W,1980,HELICOPTER-THEORY',28
  - 'SNELL-SA,1992,J-GUID-CONTROL-DYNAM,V15',25
  - 'DOYLE-JC,1989,IEEE-T-AUTOMAT-CONTR,V34',23
  - 'LANE-SH,1988,AUTOMATICA,V24',22
  - 'ISIDORI-A,1989,NONLINEAR-CONTROL-SY',20
  - 'MCRUER-D,1973,AIRCRAFT-DYNAMICS-AU',19
  - 'KWAKERNAAK-H,1972,LINEAR-OPTIMAL-CONTR',18
  - 'DOYLE-JC,1981,IEEE-T-AUTOMAT-CONTR,V26',18
  - 'MACIEJOWSKI-JM,1989,MULTIVARIABLE-FEEDBA',17
  - 'MEYER-G,1984,AUTOMATICA,V20',17
  - 'GOLDBERG-DE,1989,GENETIC-ALGORITHMS-S',17
  - 'BRYSON-AE,1975,APPLIED-OPTIMAL-CONT',17
  - 'MENON-PKA,1987,J-GUID-CONTROL-DYNAM,V10',16
  - 'MCLEAN-D,1990,AUTOMATIC-FLIGHT-CON',16
  - 'NARENDRA-KS,1990,IEEE-T-NEURAL-NETWOR,V1',16
  - 'VANDERPLAATS-GN,1984,NUMERICAL-OPTIMIZATI',15

RESULTS FROM RECENT STUDIES (CONT'D)

- **BIBLIOMETRICS - MOST CITED PAPERS - FULLERENES**

  - KRATSCHMER W 1990 NATURE V347,2773
  - KROTO HW 1985 NATURE V318,2319
  - HEBARD AF 1991 NATURE V350,1177
  - IIJIMA S 1991 NATURE V354,816
  - HEINEY PA 1991 PHYS REV LETT V66,742
  - HAUFLER RE 1990 J PHYS CHEM US V94,720
  - ALLEMAND PM 1991 J AM CHEM SOC V113,683
  - AJIE H 1990 J PHYS CHEM US V94,659
  - HADDON RC 1991 NATURE V350,602
  - KRATSCHMER W 1990 CHEM PHYS LETT V170,556
  - SAITO S 1991 PHYS REV LETT V66,527
  - KROTO HW 1991 CHEM REV V91,507
  - FLEMING RM 1991 NATURE V352,504

BIBLIOMETRICS - LESSONS LEARNED FROM RECENT STUDIES

- **VALUE OF BIBLIOMETRICS**
  - ALLOWS CRITICAL INFRASTRUCTURE IN FIELD TO BE IDENTIFIED (PROLIFIC AUTHORS/ JOURNALS/ ORGANIZATIONS)
    - ALLOWS SELECTION OF CREDIBLE EXPERTS FOR WORKSHOPS
    - ALLOWS SELECTION OF CREDIBLE EXPERTS FOR REVIEW PANELS
    - ALLOWS IDENTIFICATION OF PRODUCTIVE INDIVIDUALS AND SITES TO BE VISITED
  - ALLOWS CRITICAL INTELLECTUAL HERITAGE TO BE IDENTIFIED (HIGHLY CITED AUTHORS/ PAPERS/ JOURNALS)
  - FOR SPECIFIC AUTHORS/ PAPERS/ ORGANIZATIONS, ALLOWS PRODUCTIVITY AND IMPACT TO BE TRACKED AND ESTIMATED
  - IMPORTANT TO COMPARE ACROSS DISCIPLINES FOR PERSPECTIVE AND CONTEXT

PHRASE FREQUENCY ANALYSIS- RESULTS FROM RECENT STUDIES

- *EXAMPLES OF OUTPUT

- **COMPUTATIONAL LINGUISTICS**

  - PHRASE FREQUENCY ANALYSIS
    - IDENTIFY SINGLE, ADJACENT DOUBLE, ADJACENT TRIPLE PHRASES OF INTEREST
    - DEVELOP 'TOP-DOWN' OR 'BOTTOM-UP' TAXONOMIES IN WHICH TO GROUP PHRASES, DEPENDING ON STUDY OBJECTIVES
    - 'BIN' PHRASES AND ASSOCIATED FREQUENCIES INTO TAXONOMY CATEGORIES
    - SUM FREQUENCIES OF PHRASES IN EACH CATEGORY
    - PROVIDES ESTIMATES OF LEVELS OF EMPHASIS ON GLOBAL BASIS
      - NEEDS COMPARISON WITH REQUIREMENTS/ OPPORTUNITIES FOR CONTEXT

COMPUTATIONAL TOOLS

SELECTED PHRASE FREQUENCY EXAMPLES

AIRCRAFT-SCI DATABASE

One Word		Two Word		Three Word	
1178	AIRCRAFT	71	FLIGHT CONTROL	29	FLIGHT CONTROL SYSTEM
554	CONTROL	65	FINITE ELEMENT	19	AIRCRAFT GAS TURBINE
253	PERFORMANCE	60	CONTROL SYSTEM	15	THERMAL BARRIER COATINGS
219	HELICOPTER	40	GAS TURBINE	14	COMPUTATIONAL FLUID DYNAMICS
198	ROTOR	38	AIRCRAFT STRUCTURES	14	FINITE ELEMENT METHOD
178	COMPOSITE	38	CONTROL SYSTEMS	13	FLIGHT CONTROL SYSTEMS
176	STRUCTURES	38	HELICOPTER ROTOR	13	QUANTITATIVE FEEDBACK THEORY
154	ENGINE	37	NEURAL NETWORK	12	ANGLE OF ATTACK
149	MATERIALS	35	HANDLING QUALITIES	12	ELEMENT ALTERNATING METHOD
149	RESPONSE	30	EXPERIMENTAL DATA	12	FINITE ELEMENT ALTERNATING
146	TEST	29	CRACK GROWTH	12	HOVER AND FORWARD
143	SIMULATION	29	TRANSPORT AIRCRAFT	11	EQUATIONS OF MOTION
142	DAMAGE	27	BOUNDARY LAYER	11	FATIGUE CRACK GROWTH
140	STRUCTURAL	27	NEURAL NETWORKS	11	GAS TURBINE ENGINES
137	TECHNOLOGY	26	FLIGHT TEST	10	ELASTIC-PLASTIC FINITE ELEMENT
133	DYNAMICS	25	AIRCRAFT ENGINES	10	FLIGHT TEST DATA
127	NOISE	25	AIRCRAFT GAS	10	GAS TURBINE ENGINE
123	DYNAMIC	25	FATIGUE DAMAGE	10	MICROSTRUCTURE AND PROCESSING
123	NONLINEAR	25	FIGHTER AIRCRAFT	10	MULTIPLE SITE DAMAGE
119	AERODYNAMIC	25	FRACTURE MECHANICS	10	WIDESPREAD FATIGUE DAMAGE

PHRASE FREQUENCY ANALYSIS

- **Aircraft Strategic Taxonomy defined to group or "bin"**
  - **13 Major Categories**
  - **142 Subcategories**

- **Phrase Frequencies are summed in each subcategory a then by major category to obtain a quantitative meas of effort in area.**

- **Primarily based on 2 & 3 word phrases**

- **Useful in seeing overall trends in database related to a technologies.**

PHRASE FREQUENCY ANALYSIS
MAJOR AIRCRAFT RELATED THEMES

HIGHEST AIRCRAFT RELATED INTEREST AREAS BY MAJOR GROUPING
BASED ON PHRASE FREQUENCY ANALYSIS OF TEXT ABSTRACTS
ALSO SHOWING HIGHEST SUBCATEGORIES

(See Next Chart)

COMPARISON OF RESULTS

SCI

- **Structures**: Strength, Design/analysis, crack initiation & growth, loads & dynamics, fatigue.
- **Aeromechanics:** Aerodynamics; Design/Analysis; Performance(A/C); Drag Reduction; Wing Design; Unsteady Flow; High Lift; Wind Tunnel
- **Subsystems:** Control Systems; Neural Nets; Environmental Control Systems; Landing Gear; Subsystems (Gen.); Actuators
- **Flight Dynamics:** Stability & Control; Helicopter Rotors; Handling Qualities
- **Systems Engineering**: Fighter/Attack; Cockpit Noise; Patrol/Transport; Conceptual Design; Air Traffic Control; Airport Noise
- **Propulsion & Power**: Gas Turbine Engine; Fuels/Lubricants; Electrical Generation; Coatings; Blades/Disks; Propeller/Propfan; Electrical Power (General); Contrails
- **Avionics:** Navigation & Guidance; Decision Aids(Processing); Avionics (Gen); S/W Development; GPS; Neural Nets; Air Data; Software/Hardware(S/W)

EC

- **Aeromechanics**: Aerodynamics, Design/analysis, Performance(A/C), Wing Design, wind tunnel, drag reduction.
- **Structures:** Design/Analysis; Loads & Dynamics; Structures(Gen.); Crack Initiation & Growth; Strength; Structural Life; Aeroelastic Effects
- **Subsystems:** Control Systems; Environmental Control Systems; Neural Nets; Landing gear; Subsystems(Gen.); Fuzzy Logic; Actuators
- **Systems Engineering:** Conceptual Design; Fighter/Attack; Patrol/Transport; Air Traffic Control; Rotorcraft; UAV/UCAV; V/STOL
- **Avionics:** GPS; navigation & Guidance; Avionics(Gen.); Communication Systems; Artificial Intelligence; INS; Software/Hardware(S/W); Decision Aids(Processing); Information Management
- **Flight Dynamics:** Stability & Control; Helicopter Rotors; Handling Qualities
- **Propulsion & Power:** Gas Turbine Engine; Engines(Gen.); Electrical Power(General); Fuels/Lubricants; Electrical Generation; Blades/Disks

COMPARISON OF RESULTS (CONT'D)

SCI

- **Materials:** Composites; Metals/Alloys; NDI/NDT; Corrosion; Adhesives; Ceramics
- **Support/Logistics**: Maintenance; Take-off & Landing; Safety (Maintenance); Platform Interface; Deicing
- **Manufacturing:** Joints; Processes; Structural(Mfg); Concurrent Engineering; Composites(Mfg.)
- **Training:** Local Simulation; Manned Flight Simulation; Types(Instruction)
- **Costing:** Life Cycle Costs; Affordability of New Systems
- **Crew Systems:** Human/Machine Interface; Decision Aids; Loss of Consciousness

EC

- **Materials:** Composites; Metals/Alloys; NDI/NDT; Materials(Gen); Corrosion; Smart Materials
- **Support/Logistics:** Maintenance; Reliability; Take-off & Landing; Support/Logistics(Gen.); Runaways/Airfields
- **Crew Systems:** Displays; Decision Aids' Human/Machine Interface; Data/Information Fusion; Crew Worrkload; Cockpit
- **Manufacturing:** Processes; Composites(Mfg.); Concurrent Engineering; Joints
- **Costing:** Life Cycle Costs: Affordability of New Systems
- **Training:** Simulation(Gen.); Manned Flight Simulation; Instruction(Gen.); Distributed Simulation

PHRASE FREQUENCY ANALYSIS
MAJOR AIRCRAFT RELATED THEMES

- CATEGORY FREQUENCY NUMBERS NEED CONTEXT
- COMPARE WITH REQUIREMENTS-DRIVEN NUMBERS
- COMPARE WITH OPPORTUNITY-DRIVEN NUMBERS
  - MORE DIFFICULT TO QUANTIFY
- MORE TAXONOMY LEVELS, GREATER CATEGORY RESOLUTION, GREATER OPPORTUNITY TO IDENTIFY DEFICIENCIES/ ADEQUACIES
  - ***LABOR INTENSIVE PROCESS; NOT AUTOMATIC***

PHRASE FREQUENCY ANALYSIS LESSONS LEARNED FROM RECENT STUDIES

- **VALUE OF PHRASE FREQUENCY ANALYSIS**
  - ALLOWS LEVELS OF EMPHASIS/ EFFORT IN SPECIFIC SUBCATEGORIES TO BE ESTIMATED THROUGH 'BINNING'
  - ALLOWS JUDGEMENTS OF ADEQUACY AND DEFICIENCY IN SELECTED S&T AREAS TO BE MADE ON GLOBAL BASIS
    - NEEDS COMPARISONS TO REQUIREMENTS/ OPPORTUNITIES FOR JUDGEMENT CONTEXT
  - PROVIDES COMPREHENSIVE PICTURE OF MAJOR THRUST AREAS

PHRASE FREQUENCY ANALYSIS LESSONS LEARNED FROM RECENT STUDIES (CONT'D)

- **VALUE OF PHRASE FREQUENCY ANALYSIS** (CONT'D)
  - NO RELATIONAL INFORMATION; NOT USEFUL FOR ESTIMATING LINKAGE BETWEEN S&T AREAS
  - ***USEFUL TO APPLY TO MULTIPLE DATABASE FIELDS TO GAIN DIFFERENT PERSPECTIVES; FIELDS USED FOR DIFFERENT PURPOSES***
    - KEYWORDS
    - ABSTRACTS
    - TITLES
  - AIRCRAFT EXAMPLE
    - LONGEVITY AND MAINTENANCE IN KEYWORDS
    - NO PERFORMANCE IN KEYWORDS
    - NO TESTING IN KEYWORDS
    - OTHER AREAS SIMILAR (MATERIALS/ CONTROLS, ETC)

PHRASE PROXIMITY ANALYSIS
RESULTS FROM RECENT STUDIES

EXAMPLES OF OUTPUT

- **COMPUTATIONAL LINGUISTICS**

  – PHRASE PROXIMITY ANALYSIS
    - SELECT PHRASES OF PARTICULAR INTEREST (THEMES) FROM PHRASE FREQUENCY ANALYSIS, BASED ON STUDY OBJECTIVES
    - IDENTIFY PHRASES LOCATED PHYSICALLY CLOSE TO THE THEME PHRASES THROUGHOUT THE TEXT
    - USE NUMERICAL INDICATORS TO FILTER OUT THOSE PHRASES MOST CLOSELY ASSOCIATED WITH THEME PHRASE
      - PROVIDES ESTIMATES OF STRENGTH OF ASSOCIATION OF TEXT PHRASES TO THEME PHRASE

PHRASE PROXIMITY ANALYSIS
(EXAMPLE THEME - STRUCTURES)

Title/Block - High $I_i \geq 0.5$

Authors: Heslehurst, R.B.; Atluri, S.N.; Measures, R.M.; Brust, F.
Rubin, A.M.; Tang, D.M.; Dowell, E.H.

Journals: Journal of Solids; Journal of Intelligent Material System

Institutions: Australian Def. Force Academy; Northwestern Univ
Motoren Turbin Union Munchen GMBH; FAA Cente
Excellence in Computing.

Locations: Canberra, Australia; Munich, Germany; Moscow, Russ
Columbia, South Carolina; Atlanta Georgia; Toronto,
Evanston, Illinois; Singapore; Korea

PHRASE PROXIMITY ANALYSIS
LESSONS LEARNED FROM PHASE 1

- **VALUE OF PHRASE PROXIMITY ANALYSIS**
  - ACCESS COMPLEMENTARY LITERATURES WITH RELATED THEMES
    - HIGH POTENTIAL FOR INNOVATION AND DISCOVERY FROM OTHER DISCIPLINES
  - ALLOWS INFRASTRUCTURE (AUTHORS/ JOURNALS/ ORGANIZATIONS) RELATED TO SPECIFIC TECHNICAL AREAS TO BE IDENTIFIED
  - ALLOWS CLOSELY RELATED THEMES TO BE IDENTIFIED
    - POTENTIAL FOR IDENTIFYING "NEEDLE-IN-A-HAYSTACK"

PHRASE PROXIMITY ANALYSIS
LESSONS LEARNED FROM PHASE 1 (CONT'D)

- **VALUE OF PHRASE PROXIMITY ANALYSIS** (CONT'D)
  - ALLOWS TAXONOMIES WITH RELATIVELY INDEPENDENT CATEGORIES TO BE GENERATED USING A 'BOTTOM-UP' APPROACH
    - STARTS WITH MANY HIGH FREQUENCY THEMES
    - GROUPS RELATED THEMES INTO CATEGORIES USING PROXIMITY ANALYSIS
    - **SEE JASIS PAPER (15 APRIL 1999**) FOR DETAILED EXAMPLE OF TAXONOMY GENERATION
  - USEFUL FOR ESTIMATING LEVELS OF EMPHASIS CLOSELY ASSOCIATED WITH THE THEME

APPLICATION TO REQUIREMENTS & GUIDANCE DOCUMENTATION

- **Would it reveal major “themes” of interest?**
- **Could it be focused on relationships to “AIRCRAFT**
- **Twelve high level strategy documents were selecte**
  - **Representative of National, DOD, Navy, N-88 and N-091 policy and guidance.**
  - **All current documents available on WEB.**
  - **Prepared into single database file.**
- **Phrase Frequency Analysis performed**
- **Phrase Proximity Analysis around theme word “AIRCRAFT”.**

DOCUMENTATION

1) National Security Strategy
www.whitehouse.gov/WH/EOP/NSC/Strategy/
2) Quadrennial Review www.defenselink.mil/pubs/qdr/
3) National Military Strategy www. dtic.mil/jcs/nms/
4) Joint Vision 2010
www.dtic.mil/doctrine/jv2010/jvpub.htm
5) Joint Warfighting S&T Plan
www.dtic.mil/dstp/98_docs/jwstp/jwstp.htm
6) Defense S&T Strategy
www.dtic.mil/dstp/96_docs/strategy/strategy.htm
7) Defense Technology Area Plan
www.dtic.mil/dstp/97_docs/dtap/dtaps.htm
8) Defense Technology Objectives
www.dtic.mil/dstp/98_docs/dtos/dtos.htm
9) Forward…From the sea www.chinfo.navy.mil/navpalib/policy/
fromsea/ ffseanoc.html
10) DON 1998 Posture Statement
www.chinfo.navy.mil/navpalib/policy/
Forward…From the sea: fromsea/pos98/pos-top.html
Anytime, Anywhere
11) Forward Air Power… www.hq.navy.mil/Airwarfare/Vision/vision.htm
From the Sea
12) S&T Requirements Guidance
www.hq.navy.mil/N091/STRGCOVR.HTM

COMPUTATIONAL TOOLS
SELECTED PHRASE FREQUENCY EXAMPLES
REQUIREMENTS/GUIDANCE DOCUMENTATION

Three Word

90	COMMAND AND CONTROL
64	THEATER MISSILE DEFENSE
61	MODELING AND SIMULATION
45	WEAPONS OF MASS
39	JOINT THEATER MISSILE
37	MATERIALS AND PROCESSES
36	JOINT WARFIGHTING CAPABILITY
36	READINESS AND LOGISTICS
30	GUIDANCE AND CONTROL
30	OPERATIONS IN URBAN
29	AUTOMATIC TARGET RECOGNITION
24	BATTLE DAMAGE ASSESSMENT
24	CAPABILITY TO DETECT
24	COMBAT CASUALTY CARE
24	MANAGEMENT AND DISTRIBUTION
23	JOINT WARFIGHTING SCIENCE
23	SURVEILLANCE AND RECONNAISSANCE
21	COMMAND CONTROL COMMUNICATIONS
20	UNMANNED AERIAL VEHICLE
19	FALSE ALARM RATE
18	FOCAL PLANE ARRAY

MAJOR THEMES

THREE WORD PHRASES ONLY: CUT-OFF FREQUENCY = 5

THEME	FREQUENCY
C4/ISR	506
DETECTION & CLASSIFICATION	296
LOGISTICS/SUPPORT	231
WEAPONS OF MASS DESTRUCTION	209
JOINT WARFARE	168
THEATER MISSILE DEFENSE	157
PROPULSION	157
CONTROL SYSTEMS	111
MODELING & SIMULATION	104
MINES & MINE DETECTION	80
SIGNAL PROCESSING	54
FOCAL PLANE ARRAYS	54
AIRCRAFT	49
ELECTRICAL POWER	46
FORCE PROJECTION	43
TRAINING & REHEARSAL	42

MAJOR "AIRCRAFT" RELATED THEMES

(FREQUENCY OF OCCURRENCE)

THEME	FREQUENCY
MORE ELECTRIC A/C	120
FLIGHT CONTROL	89
LOGISTICS/SUPPORT	81
STRUCTURES	77
ROTORCRAFT DRIVE SYSTEM	65
PROPULSION	56
SUBSYSTEMS	44
V/STOL	40
ROTORCRAFT	32
AIRFIELDS	32
SELF-PROTECTION	27

CURRENT S&T PRIORITIES FOR NAVAL AVIATION

- **N88 DEVELOPED S&T PRIORITIZED CAPABILITIES. (16 NOV. 98)**

  - **57 CAPABILITIES IDENTIFIED.**
  - **DIVIDED INTO FOUR EMPHASIS AREAS**
    - * **COHERENCE**
    - * **LETHALITY/PRECISION**
    - * **SAFETY**
    - * **MECHANICAL/PROPULSION**
  - **17 OF 57 S&T PRIORITIZED CAPABILITIES ARE A/C PLATFORM RELATED.**
  - **TOP 11 GIVEN FIRST PRIORITY.**
  - **3 OF TOP 11 ARE PLATFORM RELATED.**
    - * **LONGER LIFE BEARINGS (HELO ROTORS).**
    - * **TACTICAL SITUATIONAL AWARENESS**
    - * **SAFETY OF FLIGHT.**

NAVAL AVIATION PRIORITIZED CAPABILITIES
(PLATFORM RELATED)
VS.
LEVEL OF EFFORT IN PUBLISHED LITERATURE

- **N88 Platform Related Prioritized Capabilities**
  1. - Longer Life Bearings.
  2. - Tactical Situational Awareness.
  3. - Safety of Flight
  4. - High Power Rotor Systems/Eng.
  5. - Corrosion Prevention Maintenance-A/C.
  6. - Corrosion Prevention - Detection.
  7. - Innovation Aero/Prop. - Rotorcraft
  8. - Helmet System
  9. - Wireless Sensors for Health Usage Monitoring.
  10. - Adv. A/C Control & Precision Landing.
  11. - Adv. A/C Launchers
  12. - Robotics & Automation (Deck Support)
  13. - Smart Squadron- A/C cost effective maintenance
  14. - NBC Protection
  15. - Corrosion Prevention of Support Equip.
  16. - Training & Education

- **Assessment of LOE Based on SCI & EC Da**
  1. - L
  2. - M-H
  3. - M-H
  4. - M-H
  5. - L
  6. - L
  7. - M-H
  8. - L-M
  9. - L
  10. - H
  11. - L
  12. - L
  13. - L-M
  14. - L
  15. - L
  16. - L
  17. - L

OVERVIEW

- TEXT MINING INVOLVEMENT HISTORY
- TEXT MINING DEFINITIONS
- GOALS/ OBJECTIVES/ APPLICATIONS
- TEXT MINING COMPONENTS
- BARRIERS TO TEXT MINING IMPLEMENTATION
- PILOT TEXT MINING PROGRAM
- **LESSONS LEARNED FROM PILOT PROGRAM**
- NEXT STEPS

LESSONS LEARNED FROM RECENT STUDIES

- VALUE OF/ PROBLEMS WITH TECHNICAL EXPERTS

- NEED FOR LONG-RANGE STRATEGIC PLAN

LESSONS LEARNED FROM RECENT STUDIES (CONT'D)

- ROLE OF TECHNICAL DOMAIN EXPERTS
  - CLOSE INVOLVEMENT REQUIRED IN ALL STUDY STAGES
  - ***ENHANCED EXPERT IS KEY STRATEGIC OUTPUT***
  - DATA MINING TOOLS LESS IMPORTANT THAN TECHNICAL EXPERT
  - STEEP LEARNING CURVE REQUIRED TO INTEGRATE EXPERT WITH COMPUTATIONAL TOOLS
  - SUBSTANTIAL TIME REQUIRED TO TRAIN EXPERT HOW TO USE AND INTERPRET COMPUTATIONAL TOOLS
  - LONG-RANGE INVOLVEMENT OF EXPERT WITH PROGRAM/ TOPIC AREA IS COST-EFFECTIVE BECAUSE OF LEARNING CURVE PROBLEM
  - LONG-RANGE INVOLVEMENT OF EXPERT MITIGATES AGAINST DECENTRALIZED COMPLEX DATA MINING STUDIES

LESSONS LEARNED FROM RECENT STUDIES (CONT'D)

- NEED FOR LONG-RANGE DATA MINING STRATEGIC PLAN
  - IDENTIFY ROLE OF TEXTUAL DATA MINING IN CONTEXT OF OVERALL DATA MINING
  - IDENTIFY ONR S&T DATA MINING IN CONTEXT OF NAVY S&T DATA MINING
  - IDENTIFY ROLE OF DATA MINING IN ONR BUSINESS OPERATIONS

LESSONS LEARNED FROM RECENT STUDIES (CONT'D)

- NEED FOR LONG-RANGE DATA MINING STRATEGIC PLAN (Cont'd)
  - IDENTIFIES NEEDED STUDIES AND INTEGRATION
    - DM SUPPORTS PLANNING/ REVIEWS-EVAL/ METRICS/ PR
    - OBJECTIVES
    - METRICS
    - DATA
    - EXPERTS
    - TOOLS/ TECHNIQUES
  - IDENTIFIES CRITICAL DATA TO BE GENERATED
  - (SEE: *THE SCIENTIST,* 14 SEPTEMBER 1998)
    - PRESENTLY LIMITED BY DATA AVAILABLE (EXTER/ INTER)
    - **OBJECTIVES/ METRICS SHOULD DRIVE DATA**
    - **PRESENT SITUATION IS THE REVERSE**
  - ALLOWS ECONOMIES OF SCALE FOR LARGE STUDIES; MINIMIZES DUPLICATION AND OVERLAPS OF LARGE STUDIES

OVERVIEW

- TEXT MINING INVOLVEMENT HISTORY
- TEXT MINING DEFINITIONS
- GOALS/ OBJECTIVES/ APPLICATIONS
- TEXT MINING COMPONENTS
- BARRIERS TO TEXT MINING IMPLEMENTATION
- PILOT TEXT MINING PROGRAM
- LESSONS LEARNED FROM PILOT PROGRAM
- **NEXT STEPS**

FUTURE STUDIES

RECOMMENDATIONS

- **TECHNICAL FOCUS ON MAJOR TIME AND COST DRIVERS**
  - INNOVATION AND DISCOVERY FROM COMPLEMENTARY LITERATURES
  - RETAIN QUERY COMPLEXITY; REDUCE LABOR/ TIME; EXAMINE ALTERNATE QUERIES
  - REDUCE ‘BINNING’ TIME/ LABOR; EXAMINE ALTERNATIVES
  - REDUCE TAXONOMY GENERATION TIME/ LABOR; EXAMINE ALTERNATIVE TAXONOMY GENERATORS

FUTURE STUDIES (CONT'D)

RECOMMENDATIONS (CONT'D)

- **TECHNICAL FOCUS (CONT'D)**
  - EXAMINE ALTERNATE FULL TEXT PHRASE PROXIMITY TECHNIQUES
  - EXAMINE COSTS/ BENEFITS OF
    - MULTIPLE EXPERTS
    - NUMBERS OF ITERATIONS
  - SHORTEN EXPERT LEARNING CURVES
  - EXAMINE SUPPLEMENTARY VISUALIZATION TECHNIQUES